DOCUMENT RESUME

ED 296 594 FL 017 477

AUTHOR Shohamy, Elana; Inbar, Ofra
TITLE Construct Validation of Listening Comprehension Tests: The Effect of Text and Question Type.
PUB DATE Mar 88
NOTE 35p.; Paper presented at the Annual Colloquium on Language Testing Research (10th, Urbana, IL, March 5-7, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Construct Validity; High Schools; *Language Tests; *Listening Comprehension Tests; Second Language Learning; *Test Construction; *Test Items
IDENTIFIERS *Text Types

ABSTRACT

A study examined the construct validity of second language listening comprehension tests that use different types of texts by determining to what degree the text's listenability facilitates or hinders comprehension. Three hypotheses were tested: (1) texts containing more listenable features, thus closer to the oral end of the oral/literate continuum, will yield higher scores than texts closer to the literate end; (2) second language learners will perform better on questions related to the local rather than global text elements; and (3) the interaction between oral text types and local question types will yield higher scores than the interaction between literate text types and global questions. Three text type versions of the same text (news broadcast, mini-lecture, and consultative dialogue) were used in a test administered to 150 high school seniors who were students of English as a second language. Open-ended questions concerning the texts were categorized as local, global, or trivial. Results indicated that the most literate text type, the news broadcast, was more difficult to comprehend than the other two, which contained more orally oriented features. The findings supported all three hypotheses. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED296594

CONSTRUCT VALIDATION OF LISTENING COMPREHENSION TESTS: THE EFFECT OF TEXT AND QUESTION TYPE

Elana Shohamy and Ofra Inbar
Tel Aviv University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Elana Shohamy
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A revised version of this paper was presented at the 10th annual language testing colloquium, March 5-7, 1988, Urbana, Illinois.

FL017477

BEST COPY AVAILABLE

INTRODUCTION

The definition of construct validity has been a controversial issue among psychometricians for some time. While some psychometricians consider the various types of validity, content, criterion and construct, each as constituting a seperate type of validity, Messik (1981; 1988) and Cronbach (1973), view content and criterion validity only as means for obtaining construct validity. Messik (1981, 1988) contends that content validity (obtained via experts) and criterion validity (obtained via correlations between tests) are meant to provide evidence and input for construct validity. He also adds the notion of pedagogical and instructional validity (which relate to test use and relevance), as additional sources for obtaining evidence for construct validity. Obtaining evidence that a test has construct validity is therefore a long and cubersome process that involves continuous search and research of accumulating data about a test. This type of data can be obtained from a variety of sources, such as by examining to which a test represents accurately the domain being tested, is used correctly by decision makers, is based on instruction that had taken place in school, creates biases, reflects updated theories of the trait being measured, correlates highly with other available tests considered to be good measures of the trait, or predicts successfully certain future behaviors.

Allen and Yen (1980) also view construct validation as a procedure whereby the more evidence a researcher obtains about a test, the better; and any testable prediction and hypothesis about a test should be accepted as one more piece of evidence in the long avenue of construct validation. Accordingly, there is also no one acceptable research methodology for collecting such data, and both quantitative and qualitative procedures can be used for validation purposes.

A number of hypotheses about language tests have been examined in the past few years in an attempt to accumulate data as to the existence, or lack of, construct validity. For example, Shohamy, Reves, Bejerano, (1986) researched the question of whether scores based on a single oral interaction, an interview, constitutued sufficient evidence for assigning

test takers with a valid oral test score. It should be noted that the issue was not whether the oral language trait consists of a variety of speech interaction, but rather whether the different speech interactions, known to exist in theory, manifest themselves in testing situations and will significantly affect test taker's scores on those oral tests. The results indictated that there was a need for multiple oral test interactions in order to obtain a valid representation of the oral trait. The tests which were developed based on these results required test takers to participate in a number of different types of oral interaction tasks in order to assign the test taker a valid oral proficiency score. Such a test battery could be considered more valid, it represents better the domain of oral language (content validity) but at the same time it also relates better to the theory of the oral trait.

In the area of testing reading comprehension (RC) some hypotheses about tests were examined as well. In one study the effect of the item type (multiple choice or open ended, in L-1 or in L-2) was examined, indicating that it does affect the test taker's scores on RC tests (Shohamy, 1985). Another hypothesis which was examined related to the effect of text on test taker's scores. Results of that study indicated that the type of text affects significantly the test taker's scores on RC tests (Olin, 1987). In another study (Gordon, 1987) the research question related to the strategies which test takers use in solving different types of items on RC test questions. The findings indicated that the "test strategies" which test takers use in processing information on such tests differ from "real reading".

Thus, through the testing of a variety of hypotheses about tests and by accumulating such data the researcher can obtain evidence as to the existence, or lack of, construct validity. Results of such research can then be incorporated by test developers in the construction of tests which are more valid.

The study which will be reported on in this paper is in the area of listening comprehension (LC), and the hyoptheses being examined here are about the effect of the genre used as the listening stimuli, and the type of questions, on test takers' scores on LC tests. Results of such research can provide information about the type of stimuli and the type of questions that should be included on such tests so as to increase the construct validity of LC tests.

BACKGROUND

While extensive research has been directed at RC tests, listening tends to be the "neglected skill" (McKeating, 1981; Rubin, Daly, McCrosky, 1982). Most of the studies which examine the LC skill are related to L-1 and therefore the evaluation of LC is often conducted on the basis of either reading comprehension or first language research, without reference to the issues relevant to L-2 LC (Richards, 1983).

The two comprehension skills, reading and listening, do share common features because comprehension is "always the underlying, primal purpose to any act of listening or reading" (Cambourne, 1982, p.85). They are not, however identical, and differ on a number of crucial issues, one of which is concerned with the circumstances under which the utterance is communicated: oral texts, unlike written ones, must be interpreted simultaneously while the listening is taking place. The listener is denied the option of reviewing and reconsidering the information presented, and must therefore rely on immediate comprehension, and on the ability to retain the data in the memory for further analysis. This feature affects to a great extent the content of the message and the density of propositions in the transmitted utteranes (Brown, 1978; Clark and Clark, 1977; Chafe, 1985).

Another aspect in which listening and reading differ is the form and style of the data the decoder is presented with. While oral utterances are usually regarded as context dependent, informal, unplanned and unedited, written passages are seen as manifesting the opposed features (Chafe, 1982, 1985; Ur, 1984).

Furthermore, the comprehension of the oral text, may be impeded by elements inherent in orally produced messages, such as the speaker's pronunciation, stress, intonation and rate of delivery (James, 1975; Stanley, 1978). Lexical limitation may likewise impose severe constraints on the LC of L-2 learners. James (1975) hypothesized that learners who expect to understand every single word in the spoken stimuli will despair when failing to do so, lose concentration and give up on understanding altogether. While it is true that the same process may occur in interpreting a reading text, reading provides the opportunity for re-examining the unfamiliar lexical item, and for utilizing reading comprehension strategies to infer meaning from context. L-2 LC is thus

affected by specific variables, and research findings from either RC or L-1 LC, cannot sufficiently replace direct L-2 LC research.

Despite inadequate knowledge as to what determines L-2 LC ability and performance, evaluation of that skill is conducted for decision making purposes via LC tests in and out of the classroom by major tests such as the TOEFL, RSA and the Cambridge First Certificate examination.

The question which is posed here is whether the LC stimuli are a true reflection of discoursial theories, and whether the type of questions or items which appear on LC tests, are indicative of actual comprehension, as judged by current comprehension theories.

Listening stimuli vary in the density of the oral features they contain. Thus spoken passages may range from orally-oriented texts such as dialogues, to listening stimuli which do not contain oral features, and are thus labeled only because they are transmitted through the oral medium, e.g. formal lecture (Rubin, 1980; Tannen, 1982; Rubin and Raforth, 1984). It is not known to what extent L-2 LC is affected by the density of oral features in the text, and whether achievement on a LC test which attempts to determine LC abilities or levels in L-2, are distorted by the degree of orality of the passages used as listening stimuli. This research investigated therefore the effect of this particular factor on L-2 LC test scores.

The second factor to be looked at in a comprehension test is the items used for evaluating L-2 LC abilities. L-2 LC tests are composed of various item types designed to determine the learners' comprehension. However, it is not known whether questions which tap certain comprehension skills, are better suited for tapping the LC trait. Thus, the second aspect that this research investigated is the effect of the question type on L-2 LC test scores.

FACTORS AFFECTING THE LISTENING STIMULI

Current theory rejects the concept which views text as referring solely to written language, since it is felt that texts can refer equally well to either spoken or written language samples (Halliday and Hassan, 1976). Accordingly, language is characterized by the specific linguistic or para-linguistic features it contains, (Rubin, 1980; Rubin and Raforth, 1984; Tannen, 1982). Orality and literacy are not bound by their form of presentation since written texts can be read out loud, while oral

speeches can be written and consequently read.

Texts can therefore be viewed as ranging along an oral literate continuum, where one end represents informal natural oral speech with its inherent features, and the other end represents formal literate forms of the language. Language samples are arranged according to their dominant characteristics along this spectrum. An intimate conversation, for example, will be placed on the oral end of the continuum, while a formal lecture will be found on the literate end, despite the fact that both forms of speech are delivered orally.

Among the features which determine whether a text is oriented toward the oral or the literate end of the continuum are the degree of planning; the shared context; the degree of involvement with the target audience who provides feedback and signals understanding, the prosodic features which texts embody (stress, intonation, pauses, gesures and facial expressions); the voice and clarity of delivery and text function (Tannen, 1982, 1985; Rubin, 1980, 1984, 1985; Chafe, 1982, 1985; Ricahrds, 1983; Rubin and Rafoth, 1984; Ur, 1984; Brown, 1978).

In terms of the specific linguistic elements in oral and written language, Chafe, (1982, 1985) provides a classification of the linguistic and discoursial differences between both spoken and written language. These features are manifested in elements such as idea units, the monitoring of information flow, innovations vs. conservatism, (spoken language tends to be innovative); evidentiality, (the reliability of the message conveyed); induction and deduction, (spoken language signals induction); sensory and hearsay evidence, (used more often in spoken language and conveyed via verbs and phrases such as hear, feel, it seems);

Rubin and Rafoth (1984), also mention features such as thematic redundancies which are built into spoken texts to enable listeners to cope with the fast pace and fading nature of speech and with the limits of memory, and formulaic expressions, exophoric pronoun references and ellipses, which presuppose an immediate shared context. All these, are typcial of oral texts..

Rubin and Raforth, therefore, introduce the notion of "listenability" which is a function of orality in language, or the degree to which a passage exhibits features common to oral speech. It is important to note, however, that the total list of features typical to

either modality applies only to the extreme forms of genres of spoken or written language, and the assumption that written language is more complex than spoken is too simplistic an assumption, for spoken language may be just as complex as written forms (Beaman, 1984).

It is evident that the oral/written modality is insufficient for distinguishing between texts. Thus, it is the text genres or text types, which decreed the inherent features of a particular language sample and the relative classification of a passage is to be made according to genres (Tannen, 1982; Beaman, 1984). Genres can be defined according to features such as: fields - the event in which the text is functioning; and mode - channel of presentation; rhetorical mode - narrative, persuasive, didactic, and tenor (social relations among participans) (Chafer, 1981). Other definitions of genres which relate to different variables are offered by Roeh (1982), by Crystal and Davy (1969), Joos (1967), focus on different variables in the text.

FACTORS AFFECTING THE TEST QUESTIONS

The information included in the oral stimuli is then processed through questions which attempt to tap the comprehension of the information.

Current theories view the reader/listener as being as important as the text, for the meaning does not reside only in the text but is rather a synthesis of the interaction between the reader's activated backgorund knowledge and the text. Current cognitive pscyhological views relate to schema and link it to the process of remembering and understanding. Rumelhart (1977) views the central function of schema in the construction of an interpretation of an event. The schema and the sub-schema are arranged hierarchically from general schema to its subcategories.

Van Dijk and Kintsch (1983) argue that each text type or genre has a superstructure which is the schematic representation in one's memory of variables inherent in the text overall structure, context, types of situation, degree of formality, and categories of speech and participants. This concept along with the idea of story grammar (Rumelhart, 1975) has been proven in a number of empirical studies in L-1 and in L-2. (Mandler and Johnson, 1977; Meyer and Friddle, 1984; Carrell, 1984).

In terms of processing the information, Van Dijk and Kintsch (1983)

identified two types of strategies - local, coherence strategies, which point to relations and links between the facts denoted by local clauses and sentences; and macrostrategies, which refer to the overall coherence or gist, theme or topic of a text. The local comprehension units are subordinate to the macro units in the same way that subschemata are subordinate to schema (Rumelhart, 1980).

The comprehension process can occur either by processing the information from the global macrostructure to local meaning, from schema to subschemata, in a knowledge-based manner, or from local cues to global meaning, subshemata to macrostructure schema, in a data driven manner. Knowledge-based analysis starts with hypotheses and then attempts to verify them by processing the stimuli, whereas bottom-up (or data driven) analysis starts by processing the stimuli. Both modes of text processing are interactive. The point of departure and process depends, however, on the form of analysis employed. The models mentioned in schema thory are essentially top-down, since they account for factors outside the text which influence the reading text. The theorists agree that a top down analysis is the efficient way of going about text comprehension. This indeed is the way in which skillfull readers process a text, while less able readers approach the text by interpreting local details and trying to work their way up to the overll global meaning. (Kintsch and van Dijk, 1983, Rumelhart and Ortony, 1983).

A number of studies focus on the manner in which readers deal with text processing: Meyer and Freedle (1984) showed that subjects seem to follow a certain mode of text processing whereby comprehension of main ideas preceded understanding of details, i.e., global strategies were utilized prior to local strategies in LC. Carrell (1983), however, showed that ESL readers seem to process texts by proceeding from the local elements to overall comprehension. i.e. data driven mode. Olson and Hildyard (1982) found that the efficient readers and listeners, utilized an efficient knowledge-based interactive mode of text processing, while the low level students related mostly to local details.

The modes in which readers and listeners process texts, seem to depend on variables such as being native or not, being of high or low level of proficiency as well as on additional variables such as familiarity with the topic, culture, etc. (Carrell, 1987).

Richards (1983), based on Clark and Clark (1977), relates LC theory

to schema theory, since in listening the listener also makes use of two kinds of knowledge, knowledge of syntax which enables segmentation and knowledge of the world which identifies propositions. Prior knowledge is identified in the form of schema or script in the memory of native speakers. Non-natives, however, may lack culturally specific scripts and their comprehension will therefore be impaired. Thus, the theory of LC corresponds with schema theory and we therefore relate that theory and findings to the domain of LC.

In the above review, we have shown that oral passages used as listening stimuli in EFL LC tests are comprised of various genres or text types which may be placed along a continuum according to oral factors or amount of listenability they contain. Then we indicated that the comprehension process involved in decoding an oral message consists of schema and sub-schema, thus demanding of the language learner to manipulate both global/macro and local processing strategies. It is not known, however, to what extent the elements mentioned and illustrated herewith have an effect on LC test scores. The issue here is to what degree does the use of a limited range of text types, or genres, and the type of questions used to tap comprehension of these stimuli, affect the performance of L-2 students on LC tests - do certain text types present a greater amount of difficulty than others, or to what degree does the listenability of text facilitate or hinder comprehension via listening.

Accordingly the following hypotheses were examined:

1. Texts which contain more listenable features and are thus closer to the oral end of the oral/literate continuum, will yield higher scores on LC tests of L-2 than texts closer to the literate end.

2. Learners of L-2 will perform better on questions related to the local rather than global elements of texts used as stimuli in LC tests.

3. The interaction between oral text types and local question types will yield higher scores than the interaction between literate text types and global questions used for testing LC in L-2.

METHOD

a) The Design:

The two independent variables were the listening stimuli and the tasks, and the dependent variable was the test taker's performance on the LC test. The listening stimuli consisted of six passages divided into two groups according to topics, three passages per topic. The two topics were "the common cold" and "drugs". Each topic was written in the form of three different text types or genres: a news broadcast, a lecturette and a consultive dialogue. All versions on the same topic included the same factual information.

The LC tasks were composed of eight open-ended items per topic. The items were divided into three question types: global, local and trivial, and scores on the items were to reflect the subjects' LC.

The test takers listened to two passages, each on a different topic and in a different version. The two topics were never presented in the same version, i.e. in one class the subjects may have listened to a news broadcast about the common cold and to a dialogue or lecturette about drugs. Identical questions were answered by all examinees regardless of the text type they listened to. Figure 1 presents a graphic illustration of the research design.

See Figure 1.

---

b) Description of the three text types:

1. The news broadcast was a pre-written edited monologue, delivered for the purpose of reporting novel information to unknown addressees, by a detached neutral speaker. The spoken version was identical to the written version, consisting of concise, integrated and conservative speech; literate and explicit vocabulary, no redundancies or repetitions. The sentences were of complex structure, with no pauses or fluctuations.

2. The lecturette was in a form of a monologue based on written notes delivered for the purpose of instruction or entertainment. There

was ego-involvement on the part of the speaker with references to and interaction with addressees and shared knowledge and situations with the audience. The language was less explicit, there were many repetitions, pauses and redundancies as well as monitoring of information flow. There was topic specific vocabulary using both complex and simple sentence strucutre.

3. The consultative dialoque was spontaneous, edited dialogue conducted for the purpose of consultation between an expert and an addressee, where there is constant interaction with interlocuter. Typical to this stimuli was the amount of shared knowledge and feedback, and ego-involvement on he part of both speakers. The language was implicit and fragmented, and context dependent. There were many repetitions, redudancies, interruptions, pauses and unfinished sentences. The sentences were mostly of a simple structure using colloquial innovative speech, and non-normative utterances.

The rationale for topic selection was that the test population will have some familiarity with the subjects discussed, yet the passages contained novel information about the topic. Each topic was presented in the form of three text types.

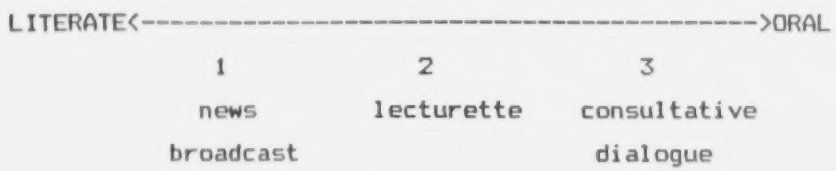
Since the same questions were used for all versions pertaining to each topic, the passages per topic had to contain the identical factual information. The passages were also controlled for lexis, so that the lexical items deemed difficult for the subjects' proficiency level appeared in all three versions of the same topic. Thus, the difference among the versions are the result of the particular discoursial features identified with each text type or genre, as mentioned in the background section.

Five native speakers of English were asked to classify the passages according to text type and results showed unanimous agreement as to the classification of the texts into the text types. Analysis of the texts based on a numerical recording of the number of times each feature appeared is displayed in Appendix A.

Figure 2, illustrates the place of the three text types chosen for the study along the oral/literaure continuum.

Figure 2

Place of Texts Along the Oral/Literate Continuum



c) Description of the question types:

1. The macro global type of questions demanded the test taker to synthesize information, to draw conclusions, to focus on cause and effect relationships, and on inferences.

2. The local questions demanded the test taker to locate details, to understand single words which have contextual support, to paraphrase and to recognize facts.

3. The trivial questions were a sub category of the "local" questions and they related to numerical details (e.g. numbers, dates, percentages) and names (e.g. of people, places), not directly related to the main topic.

The test questions consisted of 8 open ended questions per topic, on which the test takers replied in either English (L-2) or Hebrew (L-1). The answers were rated on the basis of content only, and were marked as either right or wrong. Identical questions were used for all three text types about the same topic, to enable comparison of the scores on the different text types.

The questions included macro-global, local and trivial item types. Four experts classified the questions according to the definitions provided for each question type. The questions, analyzed according to the objectives they were aimed at tapping are presented in Appendix B.

d) The sample:

The sample consisted of 150 12th grade students in six classes of an academic high school, in middle upper socioeconomic areas, in their last year of high school.

e) Data Collection:

Each class listened to a random combination of two passages dealing witn the two different topics, one on the cold and the other on drugs. The two listening passages were presented in two different text types. The listening selections were heard by the students in varying order; all the tests were adminsitered by the same external tester using the same equipment. Listening was preceded by an introduction in Hebrew as to the subject matter of the passage, in order to facilitate listening. The students were asked to read the question for each passage prior to listening; the rationale being that in real life, i.e. non-testing situations, the listener usually has a preconceived idea of the specific purpose for listening and thus focuses his listening accordingly. Each passage was played twice and subjects were allowed to take notes while listening and answer the questions during and after listening. The duration of each test was between 35 to 40 minutes.

RESULTS

In order to test the first hypothesis which stated that texts which contain more listenable features and are closer to the oral end of the continuum will yield higher scores on the LC tests, the mean and S.D. were computed for each version in each text. Table 1 presents this information for texts A and B.

See Table 1

It is evident from the results that the versions in both texts followed a similar hierarchical order of difficulty, the news being the most difficult, ($\bar{X}$=4.7708 $\bar{X}$=3.5490), followed by lecturette ($\bar{X}$=5.3265; $\bar{X}$=4.2941) and finally by the dialogue ($\bar{X}$=5.5472; $\bar{X}$=5.1042). A one way analysis of variance was computed to examine if the differences were significant. Results pointed to significant main effect at the .05 level for text B and no significant difference for text A. See table 2 below.

See Table 2

Analysis of contrasts between pairs of means was computed for both tests to examine whether there were significant differences between pairs of versions and it revealed significant differences between

versions, in both texts at the .05 level for both topics.

While no significant differences were detected between the lecturette and the dialogue, the descriptive findings point at relative difficulty compared to the other two versions, i.e. the lecturette in both texts was found to be less difficult than the news, and more difficult than the dialogue. Therefore the pattern of relative difficulty of the different text types corresponds with their position along the continuum. Figure 3 illustrates this concept.

See Figure 3

In terms of the second hypothesis which stated that the test takers will perform better on questions related to the local rather than to the global elements of texts used as listening stimuli on the LC tests, the mean scores of each of the three categories of question types were computed and compared. Findings for both tests show that the global questions presented a greater degree of difficulty than the local type questions. The third category, trivial questions, varied from text to text. While in text A it ranged the highest among the three groups, in text B they received the lowest scores and thus constituted the most difficult section of this particular test. These results are displayed in the Table 3 below.

See Table 3

A t-test analysis was conducted between the mean scores of the global and local question types, and of the local and trivial question types to exmaine whether they were significantly different. Results showed significant differences among pairs of question types at the .05 level for both texts.

Thus these findings support the hypothesis that local questions would present less difficulty than global ones. The significant differences between the question types in both texts also suggested that each question type related to different kinds of information.

Further analysis of the data on the question types was conducted in order to investigate the effect of question type on students of different levels of proficiency in order to gain insight as to the mode

of text processing employed by the test takers. The results for each text were divided into two groups according to the scores: subjects whose scores ranged from 0 to 4, were labeled "low", while test takers whose scores ranged from 5 to 8 were labeled "high". Text A was seen to include 101 "high" level test takers (67% of the total sample), and 49 "low" level examinee (33% of the sample). The "high" group for text B numbered 69 students (44.6% of the total sample), and the "low" numbered 81 (55.4%). The results of each group were classified according to question type.

Findings show that the overall pattern observed previously for each text, regarding the relative difficulty level of the question types, was maintained in each of the groups, whereby the global questions were more difficult than the local ones. The mean for question types for both groups of learners are presented in Table 4. All the items on both texts constituted more difficulty for learners in the low group than for students in the high group, and thus discriminated positively between the two levels.

See Table 4

Analysis of the responses to the local and global items in both texts showed that in the high group there were very few instances where students replied to all the global questions correctly, but failed to respond to the local questions (4% in text A, and 4.3% in text B). On the other hand, none of the learners belonging to the low group in text A succeeded in answering the global questions correctly, unless all local type items were answered appropriately. The same phenomenon reoccurred in text B, where only very few students in the low group (3.6%) supplied correct answers to global questions without having answered the local ones. Thus though both groups performed better on local rather than global, differences could be detected as to the mode utilized by each group for text processing.

Regarding the third hypothesis which stated that the interaction between oral text types and local question types will yield higher scores than the interaction between literate text types and global/macro questions used for testing L-2 LC, an analysis of covariance with repeated measures of versions and question types was employed. Results

for both texts pointed at a hierarchical pattern, whereby in almost all the interactions the highest interaction was between local question types and the dialogue, followed by the interactions between the lecturette and local questions, and finally by the news and local questions. The interaction between the global questions is lower on the ladder of mean scores, but it too followed the trend mentioned previously with the dialogue coming first followed by the other two versions. The trivial questions varied from text to text - while in text A the interaction between trivial questions and text types was the highest, in text B it was the lowest among interactions. Yet the mean scores of interaction in both cases are still arranged according to the pattern noticed previously. Table 5 presents the repeated measures of interaction between text and question type for both texts.

See Table 5

Thus, there was evidence to support this hypothesis as well. Local questions which interacted with oral text types yielded higher scores (were easier) than global questions which interacted with more literate text types. There was also a pattern of difficulty ranging among versions of text and question types in most of the interactions observed.

Contrasts between means of pairs revealed significant differences at the .05 level in text B in the local and trivial questions between the news and the other two versions (lecturette and dialogue). Results for text A likewise showed significant differences in the global questions between the news and the other two versions.

DISCUSSION

a) Summary of the Findings

The main findings of this study showed that the more literate text type, the news broadcast, was more difficult to comprehend than the other two versions which contained more orally oriented features, the lecturette and the consultive dialogue. When placed on the oral/literate continuum according to mean test score the three different versions followed the order determined by the amount of listenable features each contains, i.e. they ranged in difficulty level from the

most orally oriented version, the dialogue, followed by a less oral version, the lecturette, and finally by the most literate genre, the news. Thus text genre affected the performance of the learners in a systematic manner, depending on its place along the continuum.

In terms of the questions used to evaluate the LC performance, this pointed to the fact that subjects performed better on items referring to local cues than on those demanding the activating of global strategies. This was observed across texts as well as across versions and across levels of students. These results seem to suggest that questions classified according to different levels of cognitive processing yield various levels of performance, and that items which focus on comprehension of local information are more readily attainable than items which relate to macro overall themes in the text.

Another finding which emereged from this study was the way in which test takers of different levels processed the text. While high level L-2 listeners seemed to adopt a primarily knowledge-based approach to text processing, the low level test takers tried to abstract meaning from the message by first utilizing local cues and only then proceeded to interpret more general data which demanded strategies such as comparing, summarizing and synthesizing the information. Trivial questions were seen as affecting performance in text A and B in an utterly different manner.

CONCLUSIONS

Effect of text type: The most literate version, in both texts, the news, constituted the most difficult text type. These results can be explained by the discoursial and pragmatic features of each genre, as manifested in the text analysis conducted on the three versions.

The fact that the density of identical propositions in the news was far greater than that of the more orally oriented versions, resulted in a great number of redundant utterances in the lecturette and dialogue. The listener was able to utilize periods in which no crucial information was trnasmitted, for the purpose of internalizing and reviewing the relevant factual information he had so far heard. On the other hand, in the literate version the propositions followed one another without any pause, and the listener had to adapt to an intensive pace of data processing via comprehension staratetgies. Thus the dense and concise form of the news broadcast seems to have impeded comprehension while the

redundancies, pauses, repetitions and monitoring of information flow in the lecturette and dialogue, better enabled activating relevant strategies and achieving comprehension. The same may apply to the greater use of complex grammatical and syntactical structures such as the passive and dependent and relative clauses in the literate passage, versus the simple structure of the more orally oriented ones. This feature in the news demanded mastery in the foreign language and altogether an attentive responsive listening on a higher level than the one required by the simpler structure and style of the lecturette and dialogue.

On the pragmatic level, the dimensions identified as typical of oral versions, such as familiarity and interaction with the audience, may also have contributed to the relative ease of comprehension of these versions versus the news. The audience listening to the lecturette or to the expert in the dialogue, interacted and asked questions on aspects that the L-2 listener may have wished to clarify. As a result the speaker focussed on and re-emphasized particular issues relevant to comprehending the passage. The subjects could also identify with the listeners in the orally oriented versions, for they were of comparable age and status, and related to the matter at hand in the manner that the L-2 students may have done.

Thus it seems that the existence or non-existence of oral or literally orieneted features within the passage used as listening stimulus affected comprehension. The term "listenable" coined by Rubin and Raforth (1984), is indeed suitable for it appears that a text which contains more listenable oral features is more easily accessible to listening.

The fact that the more literate text type, the news, was found to be more difficult, may also be related to the level of familiarity of the students with the superstructures in each genre. These learners are usually exposed to a greater amount of dialogues and lecturettes than to news and it also reflects the interaction between students and teachers. Thus limited exposure to certain genres superstructures may hinder comprehension, just as ample exposure would enhance it. They may have found it more difficult to utilize their background knowledge of the specific elements which pertain to this genre, and therefore comprehension and recall were impaired.

The effect of question type: The fact that the tasks which demanded activating local strategies were found to be easier than these requiring global structures, may be explained by schema theory which distinguishes between local and global strategies employed to comprehend language samples (Rumelhart, 1980; 1983, van Dijk and Kintsch, 1983). The fact that test takers scored better on local questions, in all versions across all texts, coincides with Carrell's (1983) findings in reading comprehension. She found that the ESL subjects processed the text by first utilizing their sub-schemata (local strategies) and only then proceeding to schema type, global processing. Since the subjects in this study read the questions prior to listening, it was relatively easy to note the specific local cues required and match the given details with sub-schemata. It was apparently more difficult to generalize, infer and synthesize the given information and comply with the demands made by the global questions. Thus the majority of test takers employed a data-driven approach to text processing rather than the more efficient knowledge=-based approach utilized by native speakers (Carrell, 1983). This explanation is reinforced in the differences between the low and high levels of listeners. Low level test takers were more affected by the type of information demanded, because their performance on local and global question types differed to a greater extent than that of the learners belonging to the high group. Analysis of the answers also revealed that the low level test takers who activated their sub-schemata for text processing often did not reach the second stage of matching the over-rulling ideas and themes presented in the passage with their schema. Thus the data driven comprehension cycle was often left incomplete. Had this been a reading comprehension test, the subjects might have had the opportunity to review the passage and perhaps reach the stage of activating macro strategies, which would have possibly lead to an interaction with their own schema. It would not, therefore be sufficient to conclude that the students processed the text in a data driven and not in a knowledge based manner, for data driven processing also assumes the achievement of overall comprehension via local cues. What seems to have occured in this case is that once this mode of processing was adopted, it did not always reach the higher levels of comprehension, for while local questions were answered correctly, the global items were either ignored or answered incorrectly.

On the other hand, it can be hypothesized that subjects who employed a knowledge based approach, first made predictions about the overall theme of the passage they were about to listen to, basing them on the questions they were asked to read prior to listening. In the first listening they might have checked these hypotheses and only afterwards, in a second listening, may have filled in the information gaps by supplying the local cues. Subjects who approached the task in this manner had a better chance of completing it successfully, for rarely did students who answered the global questions correctly, failed to answer the local items, while the opposite often occured.

In terms of the trivial items, the fact that they fluctuated can be attributed to the fact that they demanded the recall of details and thus simply caused confusion. For example, the correct answer to one of the question was "one hundred billion"; however many of the listeners assumed that the numerical detail required "one hundred million" and replied incorrectly. The main conclusion here is that there are certain types of questions, those which deal with trivial information that should not be asked on LC tests. They tend to make severe demands on the test takers' memory load, and compel test takers to concentrate on memorizing insignificant details and as a result may interfere with attending to the more relevant and important tasks in the text. These results seem to suggest that questions classified according to different levels of cognitive processing yield various levels of performance, and that items which focus on comprehension of local information are more readily attainable than items which relate to macro overall themes in the text.

The interactions between text type and question types are in tune with previously discussed findings, where literate genres present an added difficulty to the L-2 learner as would a task demanding utilizing global strategies. It is logical that the combination of these two elements increased the degree of complexity for test takers.

Differences between the two texts: Text A emerged as being less difficult than text B regardless of text type. It may be speculated that the same elements which affected the degree of comprehension of the various genres could have been the reasons for the relative difficulty presented by one text as opposed to the other. Comparison of the disoursial features analyzed for the two texts does show greater

grammatical formality and complexity for text B as opposed to text A. The same holds true for text density. Thus, despite of the fact that similar versions were used for each text, the text made a difference and affected the listeners' comprehension. Another reason may be that test takers may have been more familiar with the topic discussed in text A, the common cold, than with the one in text B.

IMPLICATIONS

The results obtained in this study have direct implications to the construct validity of LC tests, since they address the issue of the extent to which the test is a good representation of the theory of the LC trait. As was noted in the introduction of the paper, obtaining evidence for construct validity requires continuous testing of hypotheses about a test. The hypotheses tested here related to the type of texts and type of questions which best tap the LC trait. The results provide information which indicate that different types of texts which are placed on different places on the oral/literate continuum, result in different performances. These findings can lead to a recommendation that tests should include a number of texts which encompass a sample of a range of genres on the continuum, with different degrees of "listenability". Judging from the results obtained here the texts which are more "listenable" are easier; although it is clear that other factors, such as familiarity with the topic, shared knowledge, etc. affect the difficulty of the text as well. Clearly, the selection of the specific type of texts from the continuum will depend on the purpose for using the tests, not all tests need to include all varieties, but certainly they should reflect these versions which best reflect the test's purposes.

Similarly, with regards to the different question types, valid LC tests should attempt to parallel the tasks which are indicative of comprehension along with comprehension theories described earlier. Thus focussing on questions which activate global strategies tasks should be designed to elicit the overall theme of the passage, to infer and generalize, and only later to relate to specific bits of local data, and utilize them to verify the predictions regarding overall ideas made before. Testing LC abilities can include local as well as global item types. Certain listening tasks, however, do demand comprehension of

relevant details, and it should not be conceived that usage of global items only would suffice for the purpose of evaluating LC. Trivial questions, which related to precise recall of names or numerial data, were found to be unstable and serve no meaningful purpose as evaluation tools, and should therefore be avoided altogether.

Finally, it is recommended that LC tests should present a variety of listening stimuli, a variety of tasks and a variety of interactions, to better reflect the trait of LC and result in more construct valid LC tests.

BIBLIOGRAPHY

Allen M. and Yen W. 1979. Introducing Measurement Theory, Brooks/Cole Publishing Company Monterey, California.

Beaman, K. 1984. Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In Coherence in Spoken and Written Discourse, Tannen, D. Eds. Norwood N.L.: Ablex.

Brown, G. 1978. Listening comprehension, In The Teaching of Comprehension. London: The British Council English Information Center.

Cambourne, B. 1982. Oral and written relationship: A reading perspective. In Exploring Speaking and Writing Relationship, Connections and Contrasts. Knoll, B.M. and Vann R.J. (eds.) Urbana III. The National Council of Teachers of English.

Carrell, P. 1983. Three components of background knowledge in reading comprehension. Language Learning, 33 (20), 183-208.

Carrell, P. 1984. Evidence of a formal schema in second langauge comprehension. Language Learning, 34(2), 87-112.

Chafe, W. 1982. Integration and involvement in speaking, writing and oral literature. In Spoken and Written Language: Exploring Orality and Literacy, Tannen, D. (Eds.) Norwood N.J.: Ablex

Chafe, W. 1985. Linguistic differences produced by differences between speaking and writing. In Literacy Language and Learning, Olson, D. Torrance, Nancy and Hildyard, Angela (Eds.) Cambridge: CUP.

Clark,H. and Clark E. 1977. Psychology and Language. NY.: Harcourt, Brace Jonovich, Inc.

Cronbach, L. J., 1971. Test validation. In R.L. Thorndike (Ed.) Educational Measurement (2nd ed.) Washington. DC: American Council on Education.

Crystal D. and Davy D. 1969. Investigating English Style. London: Longam Group, Ltd.

Gordon, C. 1987. The Effects of Testing Method on Achievement in Reading Comprehension Tests in English as a Foreign Language. MA Thesis. Tel Aviv Univeristy.

Hildyard, A. and Olson, D. 1982. On the comprehension and memory of oral versus written discourse. In Spoken and Written Language, Tannen, D. (Eds.) Norwood, N.J.: Ablex Publishing Co. 19-32.

James, K., 1975. Notetaking in lectures: Problems and strategies in English for Special Purposes, Cowie, A.P. and Heaton, J.B. (Eds.) Univeresity of Reading, England.

Joos M. The Five Clocks, New York. Harcourt Brace and World.

Mandler, J. and Johnson, N. 1977. A code in the node: The use of a story schema in the retrieval. Discourse Processes, 1, 14-35.

McKeating D. 1981. Comprehension of listening in English as an International Language. In The Teaching of English as International Language. In Abbott G. and Wingrad P. editors. Glasgow. Williams Collins and Sons. Ltd.

Messick, S. 1981. Evidence and ethics in the evaluation of tests. Educational Researcher, 10, 9-20.

Messick, S. 1988. Meaning and Values in Test Validation; the Science and Ethics of Assessment. Paper presented at the AERA Convention, New Olreans.

Meyer, B. and Freedle, R. 1984. Effects of discourse type on recall. American Educational Research Journal, 21,141-143.

Olin, R. 1987. The Effect of Connectors on Reading Comprehension. M.A. Thesis, Tel Aviv University.

Richards, J. 1983. Listening comprehension - approach, design, proccdure, TESOL Quarterly, 17, 189-220.

Roeh, I. 1982. Rhetorics of News on Isreali Radio. Studiennerlog. Doctor N. Brockneyer, Bochum.

Rubin, D. and Raforth, B. 1984. Oral language: a criterion for selecting listenable materials. (Unpublished paper).

Rubin D. 1980. Theoretical taxonomy of the difference between oral and written language. In Theoretical Issues in Reading Comprehenison.
Rubin, D. and Daly J. McKorsky J. and Mead 1984. A review and critique of procedures for assessing speaking and listening skills among preschool through grade twelve students. Communication Education, 34, 285-303.

Rumelhart, D. 1975. Notes on schema for studies. In Representation and Understanding, Studies in Cognitive Science. Barrow, Daniel and Collins, Allan (Eds.) N.Y. Academic Press Inc. 211-236.

Rumelhart, D. and Ortonoy, A. 1977. The representation of knowledge in memory, In Schooling and the Acquistion of Knowledge, Anderson, R. Spiron, R. and Montague W. (Eds.) Hillside N.J. E. Erlbaum Associates, 99-135.

Rumelhart, D. 1980. Shcemata: The building blocks of cognition. In Theoretical Issues in Reading Comprehenion.

Schafer J. 1981. The linguistic analysis of spoken and written texts. In Exploring Speaking and Writing Relationship, Connections and Contrasts. Knoll, B.M. and Vann R.J. (eds.) Urbana III. The National Council of Teachers of English.

Shohamy, E. 1985.Does the testing make a difference? The case of reading comoprehension. Language Testing, 1, 2.

Shohamy, E. Reves, T. and Bejerano, Y. 1986. Introducing a new comprehensive test of oral proficiency. English Language Teaching Journal, 49, 212-220.

Tannen, D. 1982. The oral literate continuum of discourse. In Spoken and Written Language. Tannen, D. (Ed.) Norwood, N.J.; Ablex, 1-6.

Tannen, D. 1985. Relative focus of invovlement in oral and written discourse. In Literary Language and Learning. Olson D. Torrance, Nancy and Hildyard, (Eds.) Cambridge: CUP.

van Dijk, T. and Kintsh W. 1983. Strategies of Discourse Comprehension, N.Y. Academic Press.

Ur P. 1984. Teaching Listening Comprehension. Cambridge. CUP.

TABLE 1 - The Mean and SD of the Three Versions in Texts A and B

text	version	n	mean	SD
Common Cold	news	48	4.7708	1.7165
	lecturette	49	5.3265	1.4633
	dialogue	53	5.5472	1.7712
Drugs	news	51	3.5490	1.8365
	lecturette	51	4.2941	1.6161
	dialogue	48	5.1042	1.7165

TABLE 2 - A One-Way Analysis of Variance for Texts A and B

SOURCE	DF	Sum of Squares		Mean Squares		F Ratio		F Prob	
		Text A	Text B	Text A	Text B	A	B	A	B
Between groups	2	15.9066	59.80511	7.9533	29.9026	2.891	10.043	.0587	.000*
Within groups	147	404.3868	437.6949	2.7509	2.9775				
Total	149	420.2933	497.5000						

TABLE 3 - The Mean and SD of the Question Types in Texts A and B

Quest.Type	Text A ("Common Cold")		Text B ("Drugs)	
	Mean	SD	Mean &	SD
Global	46.89	.3163	47.78	.3463
Local	66.67	.2818	77.00	.3206
Trivial	82.89	.2404	44.22	.2877
	n = 150			

TABLE 4 - Classification of Scores obtained by "High" and "Low" levels of EFL students According to Global and Local Questions for Texts A and B

	Text A		Text B	
Level	High	Low	High	Low
Que.Type	mean(%)	mean(%)	(mean(%)	mean(%)
Global	64	18	80	30
Local	77	45	96	52
	n = 101	n = 49	n = 69	n = 81

TABLE 5 - Repeated measures of interaction between text and question type for texts A ("Common Cold) and B ("Drugs)

Version / Question Type	News A %	News B %	Lecturette A %	Lecturette B %	Dialogue A %	Dialogue B %
Global	37.438	45.049	50.367	42.471	52.151	56.250
Local	61.224	60.784	64.583	84.294	73.585	85.417
Trivial	78.583	32.647	86.490	43.765	83.717	57.000

Figure 1:

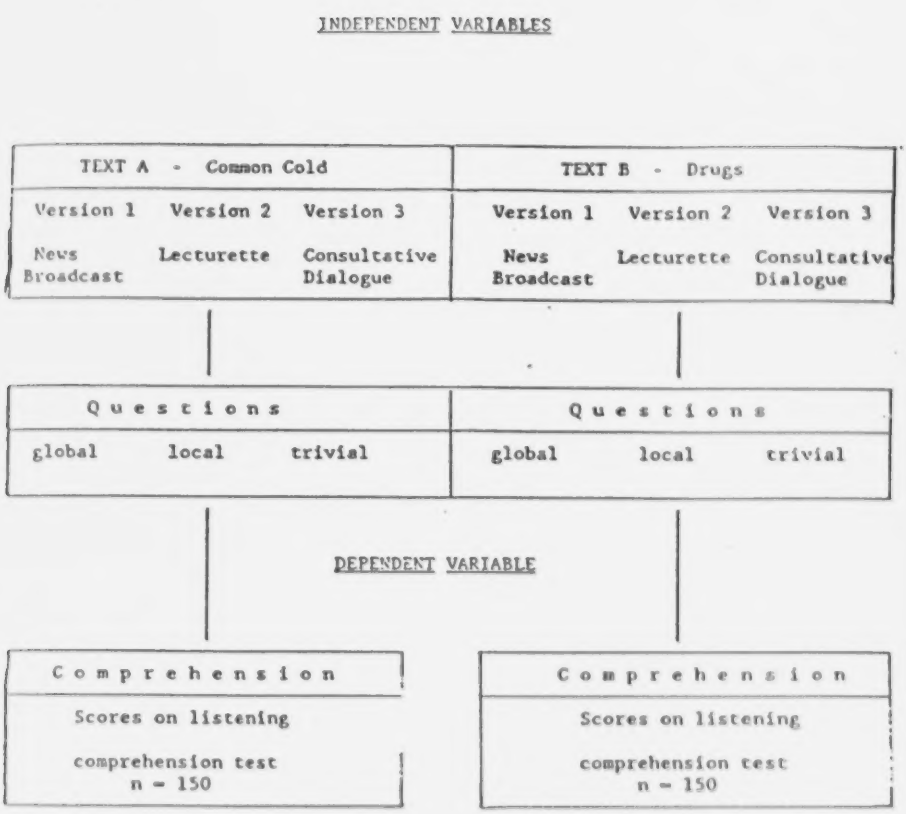


Figure 1: A graphic representation of the research design used in order to examine the hypotheses of this study.

Figure 3: Text A ("Common Cold")

	news broadcast	lecturette	dialogue	
LITERATE	($\bar{x}$ = 4.7708)	($\bar{x}$ = 5.3265)	($\bar{x}$ = 5.5472)	ORAL

Text B ("Drugs")

	news broadcast	lecturette	dialogue	
LITERATE	($\bar{x}$ = 3.5490)	($\bar{x}$ = 4.2941)	($\bar{x}$ = 5.1042)	ORAL

Appendix A.

Text analysis According to oral/written discourse features (based on Chafe, 1982, 55)

Text Types / FEATURES	news cold	news drugs	lecturette cold	lecturette drugs	dialogue cold	dialogue drugs
1. length (in words)	about 180	about 250	about 500	about 650	about 630	about 800
2. ego involvement	-	-	20	14	26	31
3. audience involvement	-	-	22	16	38	28
4. sub.matter involv.	-	-	2	2	6	4
5. sensory evidence	-	-	3	2	3	4
6. hearsay evidence	-	-	2	1	2	2
7. conversation fillers	-	-	10	7	15	14
8. monitoring info.	-	-	24	14	34	35
9. false starts (unfinished sentences)	-	-	3	3	6	12
10. repetitions	-	-	5	3	8	8
11. colloquial lang.	-	-	-	-	4	4
12. contractions	-	-	8	10	25	23
13. use of this/that to refer to past state or event	-	-	6	8	9	12
14. use of "che"	11	12	8	8	4	10
15. use of the passive	4	6	6	5	1	1
16. relative clauses	2	2	7	3	4	6
17. complement clauses	5	5	8	5	6	8
18. dependent clauses introduced by subordinating conjunctions	7	4	4	3	-	2
19. nominalization	6	2	2	-	1	1
20. coordinating constructions	2	2	4	4	6	8
21. postposed past participles	4	3	-	1	1	2
22. inductive reasoning	-	-	3	2	3	4
23. sloppy grammar (includes #13)	-	-	-	-	11	15
24. text density (no. of identical propositions per number of words)	14/80	22/250	14/500	22/650	14/630	22/800

Note: Since an extreme oral form was not included in the listening stimuli certain features which typify such discourse are missing from the analysis, as are others which do not appear in the texts.

Appendix B Text analysis according to oral/written dimensions

Dimensions \ Texts	news (cold)	news (drugs)	lecturette (cold)	Lecturette (drugs)	dialogue (cold)	dialogue (drugs)
1. spontaneity	0	0	1	1	2	2
2. shared knowledge	1	1	2	2	2	2
3. size of audience	0	0	1	1	3	3
4. physical presence of audience	0	0	1	1	1	1
5. feedback	0	0	2	2	3	3
6. familiarity with audience	0	0	2	0	2	2
Total	1	1	9	7	13	13